THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, *15*(3), 233–250

Auditory Alert Characteristics: A Survey of Pilot Views

Guy Peryer, Jan Noyes, Kit Pleydell-Pearce
Department of Experimental Psychology
University of Bristol

Nick Lieven
Department of Aerospace Engineering
University of Bristol

This survey obtained feedback from 50 commercial pilots on the design characteristics of a set of auditory flight deck alerts. The first section in the online questionnaire assessed pilot views on current loudness levels and prevalence of cognitive impairment from alert presentation. The second section in the questionnaire required the pilots to rate the effectiveness of 10 auditory alerts currently in operation on a large commercial aircraft. In general, the alerts were rated favorably on informational content variables, but unfavorably on presentation style variables. Loud, continuous alerts can distract pilot attention away from processing task-related information, increasing the potential for error. New methods for alert presentation are suggested, based on real-time analysis of flight deck noise. This is argued to offer a more suitable environment for problem solving, decision making, and communication during alert situations.

Flight deck alerts, designed to be pilot aids, are often perceived as being intrusive and aversive (Sorkin, 1988). Unpopular features of alert systems include high numbers of false alarms (Bliss, 2003) and undesirable presentation characteristics (Mårtensson, 1995; Patterson, 1982). Both reduce the credibility of the alert systems. This article focuses solely on the acoustic characteristics of alert signals. The motivation for carrying out this research was to obtain feedback from pilots on current alert characteristics and to assimilate information con-

Requests for reprints should be sent to Jan Noyes, University of Bristol, Department of Experimental Psychology, 8 Woodland Road, Bristol, BS8 1TN, England. Email: j.noyes@bristol.ac.uk

cerning a pilot's cognitive capabilities and limitations during alert events (Noyes & Starr, 2000). It is important to continue to run pilot feedback studies to see how well research guidelines have infiltrated into operational practice and to report suggested changes.

A primary role of flight deck auditory alerts is to direct the pilot's attention to a particular system event. In a three-stage process the pilot must recognize the alert, analyze its meaning in relation to other information presented on the flight deck, and aim to resolve any abnormal or unsafe operating conditions (Hawkins, 1993; Pritchett, 2001). Consequently, auditory alerts represent an acoustic cue to possible complex problem-solving and decision-making tasks. The way a pilot is alerted to an abnormal system state will influence subsequent task performance following alert detection (Wickens & Hollands, 2000). If an auditory alert acts as a stressor, the immediate inclination for pilots is to seek a means of silencing the noise rather than analyzing the meaning of the alert and resolving any unsafe operating conditions. In these circumstances, pilots view auditory alerts as a source of irritation (Patterson, 1982). Furthermore, increasing pilot stress levels during warning events may lead to an increased potential for human error (Rigby & Edelman, 1968; Wickens & Hollands, 2000). Designers need to ensure that alerts have appropriate characteristics that enable swift, accurate understanding and do not prevent flight deck communication or impair the cognitive functioning of the crew. This was summarized in a Federal Aviation Administration (FAA) Advisory (2000): "Warning information must be provided to alert crew to unsafe operating conditions, and enable them to take appropriate corrective action. Systems, controls, and associated monitoring and warning means must be designed to minimize crew errors which could create additional hazards" (p. 126).

ALERT INTENSITY

The most common complaint from pilots concerning auditory alerts is excessive volume (James, 1996; Patterson, 1982). For readers who are unfamiliar with the previous literature, a summary is given here.

In a Lufthansa survey (Uchtdorf & Heldt, 1989) of the Boeing 737–300 and the Airbus 310–200 flight deck systems, 98% of the pilots regarded auditory alerts as extremely important to flight safety. However, almost all participants complained that the alerts were presented at an excessive volume. Manufacturers tend to take the "better safe than sorry" approach when setting the alert intensity level (James, 1996). This approach is understandable, as an alert that is masked by flight deck noise will go undetected and render the auditory warning system useless. However, high-intensity sounds have been reported to have undesirable physiological and psychological effects.

Adverse Effects From the Onset of a High-Intensity Alert (Auditory Startle)

The human auditory startle response can be elicited from sounds that have abrupt onsets and high sound pressure levels (Björk, 1999; Thackray & Touchstone, 1970; Turpin, Schaefer, & Boucsein, 1999; Wilkins, 1982). There are various physiological changes that accompany the startle response: chemical changes in the blood (e.g., adrenalin), changes in heart rate and blood pressure, changes in breathing and sweat gland activity, and brief changes in skeletal muscle tension (Wilkins, 1982). The startle response itself is thought not to endanger health; however, deficits in performance can be observed for a short period after the startle response has occurred. Thackray and Touchstone (1970) used a tracking task as a performance measure. A burst of white noise of over 100dB occurred unexpectedly in a 30-min testing session. Maximum disruption to tracking task performance was observed in the first 2 sec following startle, showing that perceptual motor recovery is relatively fast. However, significant disruption was observed up to 10 sec following startle.

Pilot reports of cognitive impairment from excessively loud alerts can be accessed via the Aviation Safety Reporting System (ASRS) database. The reports are written in shorthand, submitted voluntarily, and are anonymous. Within the database there are numerous examples of how loud alerts cause disorientation, confusion, and occasionally unintended actions. Moreover, nearly all of the reports that mention loud alerts claim that the alert directly contributed to the breach in flight safety. For example, "the aspect of this incident that I must comment on is how distracting and potentially disorienting a large sudden change in noise level … can be. When these large sudden changes occur, the crew can be momentarily immobilized by a startle reaction" (National Safety Data Analysis Center [NASDAC] Report No. 260299, 1993). In this situation a pilot might not be able to recognize anything beyond the presence of an alert signal and all information regarding alert meaning and corrective action is temporarily lost (Pritchett, 2001). Perceptual characteristics that increase the startle response include a more abrupt onset, otherwise known as the attack or rise time of a stimulus (Björk, 1999; Turpin et al., 1999); higher intensities (Björk, 1999; Turpin et al., 1999); unpleasant signals (Bradley & Lang, 2000); and darkness (Grillon, Pellowski, Merikangas, & Davis, 1997).

Adverse Effects From the Duration of a High-Intensity Alert

Some flight deck auditory alerts continue until corrective action takes place. Continuous alerts of high intensity have been reported to be very distracting. As an example, "While attempting to inhibit this loud continuous distracting aural warning, I inadvertently activated the alternate LNDG gear extension switch while trying to activate the GPWS gear override. There was then a very loud roaring noise in the cockpit further aggravating our anxiety" (NASDAC Report No. 412066, 1998). A review of distraction caused by background sounds is contained in Banbury, Macken, Tremblay, and Jones (2001), which concluded that both speech and nonspeech sounds have a distracting effect on serial recall and tasks that involve seriation (the process of placing items in order). Levels of distraction have been documented as being independent of intensity; however, as yet, no studies have examined auditory distraction using flight deck presentation levels (i.e., above 85dB). It is hypothesized that higher intensity alerts would have a qualitative effect on distraction (Banbury et al., 2001) purely due to the stressful nature of experiencing loud sounds. The auditory modality is very sensitive to change, which makes it ideal for warning presentation; however, if the alert continues, attention, even at a rudimentary level, is still devoted to processing the auditory signal and is diverted away from processing task-related information (Banbury et al., 2001; Wickens & Hollands, 2000). A secondary adverse effect from loud and continuous auditory alerts is the masking of communications (Patterson, 1982; Pritchett, 2001). For example, "the B737–900 has an alt alert warning audio set too high and interferes with ATC coms … I consider this hazardous to safety, and it adds to freq congestion when missed ATC instructions must be repeated" (NASDAC Report No. 543414, 2002).

It is essential that alert signals are detected and that the flight crew are made aware that a problem is ongoing. However, care needs to be taken when setting intensity levels and alert durations. Auditory startle, task distraction, and communications masking are all issues that increase the chance of pilot error.

CURRENT GUIDELINES FOR ALERT DESIGN

Alert Intensity

Patterson (1982) suggested a model of alert presentation based on the intensity level and spectral content of background noise. Alerts should be presented at least 15dB above the threshold of alert detection and limited to 30dB above threshold to prevent task disruption. The guidelines put forward by Patterson are widely accepted and have been incorporated into the U.K. Ministry of Defence Standard (INT DEF STAN 00–25, 2004). In addition, similar guidelines can be found for the International Organization of Standardization (ISO 7731, 1986) and the U.S. Human Engineering Standard (MIL–STD–1427F, 1999). Loudness levels of alerts are referred to specifically in this survey as a means of discovering if alert intensity is still a problem 20 years on from Patterson's study.

Alert Comprehension: Distinctiveness and Associability Guidelines

The U.S. Human Engineering Standard (MIL–STD–1427F, 1999) suggested that the first 0.5 sec of an auditory alert signal that requires immediate action should be distinctive from the first 0.5 sec of any other alert signal that may occur at that time. This reduces the need for the pilot to search visual displays for alert verification. Distinctiveness as an alert characteristic is central to effective design and is referred to specifically within this survey. An example of a highly distinctive alert that is easily associated with its meaning is the fire bell. Patterson's (1982) study indicated that on 98% of trials the fire bell was recalled correctly and not confused with any other alert sound. This is arguably due to an increased familiarity with fire bells outside the flight deck environment. However, this does demonstrate how nonverbal auditory signals can be highly effective alerts, providing there is a sufficient association between the sound and the alert meaning (Ulfvengren, 2003).

As the need for more auditory alerts increases, it seems logical to use more speech signals on the flight deck (Doll & Folds, 1986). The advantages of speech signals are that they are reliable and versatile and require no additional learning. A disadvantage is that there is already a lot of speech on the flight deck and a lack of perceptual contrast may reduce the effectiveness of a speech alert (Patterson, 1982). A compromise advocated by numerous researchers is to have a nonspeech signal followed by a short speech signal. This way, the nonspeech signal captures the pilot's attention and the speech signal confirms the alert meaning. Preference of alert type is referred to specifically in this survey.

Urgency Mapping

Flight deck visual alerts are generally prioritized into categories of urgency: warnings, cautions, and advisories. Warning messages appear in red on visual displays and require immediate crew action. Caution messages appear in amber and require immediate crew awareness. Both warning and caution messages are accompanied by an auditory alert.

Alerts tend to occur in phases of flight with the highest workload (Woods, 1995). Consequently, when an alert is triggered, pilots are usually faced with a situation in which they have to coordinate subsequent action based on the situational urgency of the alert message. More specifically,

> Interruptions and distractions create the demand to suspend ongoing activities, defer planned actions, and remember to resume suspended actions and/or execute intentions in the future. In doing so, they constantly threaten to sidetrack pilots and make

> them vulnerable to a variety of potential errors. (Loukopoulos, Dismukes, & Barshi, 2001, p. 2)

A guideline that has developed out of recent research is *urgency mapping* (Edworthy, 1994), in which the urgency of the event matches the perceived urgency of the auditory alert. This is important so that cautions are not designed with higher urgency characteristics than warnings. Various studies have attempted to find factors that influence perceived urgency of auditory alerts (Burt, Bartolome, Burdette, & Comstock, 1995; Haas & Casali, 1995; Haas & Edworthy, 1996; Hellier & Edworthy, 1999; Momtahan, Hétu, & Tansley 1993). Collectively, research indicates that perceived urgency can be manipulated using intensity levels, temporal characteristics, and spectral characteristics. For example, auditory alert urgency can be increased by raising intensity level, decreasing interpulse interval, and raising pitch. A case study on the urgency mappings of alerts on the Griffon CH–146 helicopter concluded that the acoustic characteristics of the tested alerts did not adequately map onto levels of situational urgency (Arrabito, Mondor, & Kent, 2004). If urgency mismatching does occur, laboratory research indicates that task performance is undermined (Edworthy, Hellier, Walters, Weedon, & Adams, 2000). The level of urgency mapping for the selection of alerts is assessed within this survey by correlating event urgency with the perceived urgency of the alert sound.

METHOD

Participants

A within-subjects questionnaire design was employed. Initially, a pilot study was carried out to refine the content of the questionnaire and test the technical operations of Web-based data collection. Sixteen members of the University Air Squadron participated. All individuals had flying experience, although some had relatively few solo hours.

The majority of professional pilots who participated were contacted via the Bluecoat organization. The remainder of the participants were British Airways pilots. Table 1 lists the demographic information of the 50 pilots who took part in the survey.

Survey Design

Section A assessed the subjective opinions of pilots on current loudness levels, prevalence of startle and cognitive impairment as a result of auditory alert presentation, preferences of alert type, and the level of urgency of a list of warning events. The questions are listed here. At every stage participants were encouraged to give further comments.

TABLE 1
Participant Details

Demographic Factors[a]	*M*	*Mode*
Age	44.5	39
Flying hours	9,922	15,000
Years on current aircraft type	4.99	5
Familiarity with auditory alerts[b]	6.7	7
Display Technology on Current Aircraft Type[a]	*Frequency*	
Analogue	2	
Hybrid (analogue/digital)	13	
Glass cockpit	35	

[a]$n = 50$. [b]Based on a 7-point scale ranging from 1 (*unfamiliar*) to 7 (*very familiar*).

1. In general, how would you describe the loudness levels of auditory alerts on the flight deck? Scale ranged from 1 (*far too loud*) to 7 (*far too quiet*).
2. Have you ever experienced an auditory alert that has impaired your performance in a high-workload situation? (Yes/No. If Yes, please give further information.)
3. Have you ever been startled by an auditory alert on the flight deck? (Yes/No. If Yes, please give further explanation.)
4. Do you prefer flight deck auditory alerts with (a) speech elements, (b) nonspeech elements, or (c) a mixture of the two? Please give a reason for your choice.
5. Please rate how urgent the situation is if the following auditory alert is triggered? (fire, traffic, windshear, system caution, configuration, ground proximity, autopilot disconnect, overspeed, decision height, 500 ft). Scale ranged from 1 (*not at all urgent*) to 7 (*critical*).

Section B assessed the subjective opinions of pilots on the effectiveness of a set of 10 alerts currently in operation on a large commercial aircraft. These included 6 nonspeech alerts (fire, configuration, autopilot disconnect, system caution, decision height, overspeed), 2 speech alerts (traffic, 500 ft), and 2 combination alerts that had both speech and nonspeech components (ground proximity and windshear). All alerts were recorded in quiet conditions. From the literature a checklist of desirable characteristics was assembled to assess auditory alert efficiency. These included the following:

1. Detection: Auditory alerts need to be presented at a sufficient level above ambient noise to ensure detection but limited to levels that do not cause auditory startle and task disruption.

2. Sound type: The type of sound chosen for an alert should be highly distinctive and easily associable with the alert function.
3. Urgency: The perceived urgency of the alert needs to map the criticality of the situation.
4. Duration: The alert meaning should be delivered succinctly. Auditory alerts should not prevent fluent flight deck communication or disrupt piloting tasks.

A sound file was embedded at the top of each page. The participant was instructed to listen to the alert sound several times before proceeding with the questions listed next. Ratings were given on a 7-point Likert scale with a midpoint of 4. The questions were structured so that desirable characteristics (e.g., an alert that is highly distinctive) would receive a rating of greater than 4 and undesirable characteristics (e.g., an alert that has a high likelihood of startle) would receive a rating lower than 4.

1. How familiar are you with this sound? Scale ranged from 1 (*never heard*) to 7 (*very familiar*).
2. How distinctive is this sound from other auditory alerts? Scale ranged from 1 (*very similar to others*) to 7 (*very distinctive*).
3. How easy would it be to detect this sound in a high-workload situation? Scale ranged from 1 (*very hard*) to 7 (*very easy*).
4. How easy is it for you to associate this sound with the alerting event? Scale ranged from 1 (*very difficult*) to 7 (*very easy*).
5. What is the likelihood of you being startled by this sound? Scale ranged from 1 (*very likely*) to 7 (*very unlikely*).
6. If this sound was repeated several times or presented continuously, how irritating would it be? Scale ranged from 1 (*very irritating*) to 7(*not at all irritating*).
7. If this sound was repeated several times or presented continuously, how distracting would it be? Scale ranged from 1 (*very distracting*) to 7 (*not at all distracting*).
8. What level of urgency does the sound itself convey? Scale ranged from 1 (*very low*) to 7 (*very high*).
9. Overall, how effective is this sound as a flight deck alert? Scale ranged from 1 (*very ineffective*) to 7 (*very effective*).

RESULTS

Section A: Loudness Levels, Performance Impairment, and Alert Preferences

1. In general how would you describe the loudness levels of auditory alerts on the flight deck? (see Figure 1; based on a scale ranging from 1 [*far too loud*] to 7 [*far too quiet*]).

Twenty-six participants (52%) gave a rating of lower than 4 (the midpoint) and the mean rating was slightly above 3 (3.16), indicating that current alert loudness levels are regarded as too loud by over half of the pilots tested. The means are given to permit comparisons with other research findings, notwithstanding their doubtful statistical use with pseudo-interval data. In addition, a series of pilot comments have been selected that highlight design issues.

Comment: "Consideration of volume levels with Active Noise Reduction headsets needs to be made. Sometimes it seems that the warnings and alerts have all been turned up way too much and are now deafening, leading to a tendency to try to block them out. Practicing ignoring them surely must be dangerous."

2. Have you ever experienced an auditory alert that has impaired your performance in a high-workload situation? (Yes/No. If yes, please give further information.)

Approximately half of the participants (46%) had experienced alerts that impaired their performance. Analysis of the comments offered in response to this question reiterates previous research in that alerts have a tendency to mask communications and dominate attention.

Comment: "Critical yet nonaural warnings (e.g., approaching altitude alert light) become invisible when competing with multiple, nuisance (and very loud) traffic alerts."

3. Have you ever been startled by an auditory alert on the flight deck? (Yes/No. If yes, please give further information.)

A high proportion of participants (74%) had experienced a startle response from auditory alerts on the flight deck. Analysis of the additional comments offered by the participants revealed that excessively loud alerts cause a disruption to piloting tasks and increase stress levels on the flight deck.

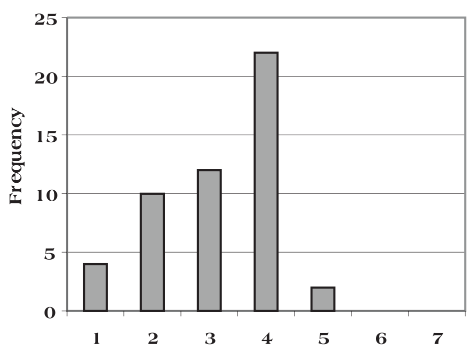


FIGURE 1 Alert loudness ratings. Based on a scale ranging from 1 (*far too loud*) to 7 (*far too quiet*). $N = 50$. Loudness levels: $M = 3.16$, Mode = 4.00, $SD = 1.06$.

Comment: "When the audible caution/warning is so loud it startles me, it just adds more distraction/confusion when I probably have more than I need. Silencing the loud noise temporarily takes precedence over (1) good flying, (2) analyzing the problem, and (3) solving the problem."

4. Do you prefer flight deck auditory alerts with speech elements, nonspeech elements, or a combination of the two?

This question examined which type of alerts pilots find most useful. The majority of pilots (64%) responded in favor of combination alerts. Analysis of the comments indicated that nonspeech sounds are beneficial in alerting the pilot to a fault, but an added spoken message provides verification of alert meaning.

Comment: "Specificity when sounding an alarm reduces the discernment time for a pilot. When a nonspeech alert is sounded, it takes time to scan the cockpit and recognize the source. Even familiar bells and chimes require absolute verification. This is not simply because the same alert may be used to designate different alarm conditions, flight regime depending … it's also because pilots who jump from airplane to airplane (as I have) may not have the desired familiarity or confidence in discerning the meaning of a particular chime or horn."

Section B: Alert Effectiveness Ratings

The 7-point Likert scale had a midpoint of 4. Means of 4 and higher represented desirable alert characteristics and means lower than 4 represented undesirable alert characteristics. All the alert sounds scored highly on the effectiveness variable with the exception of the overspeed alert and the decision height alert (see Appendix). The overall effectiveness scores for these two alerts were surprisingly low. However, this may be explained by the low familiarity ratings on both alerts. Other noteworthy results were observed on the startle, irritation, and distraction variables. Alerts that were rated as likely to cause startle were the fire bell, autopilot disconnect, configuration siren, system caution, ground proximity, and windshear alerts. If any of the alerts were presented repetitively or continuously they were rated as being irritating and distracting (see Appendix). All the alerts received a meaning rating of lower than 4 on these two variables, indicating a negative result. The mean, mode, and median values for the whole set of alerts are given in Table 2.

A principal component analysis (PCA) was performed on the data to detect structure in the relationship between rating variables. The alert urgency ratings were omitted from this analysis because they were only deemed relevant to the urgency mapping analysis, mentioned in greater detail later. Prior to performing the PCA, the suitability of data for factor analysis was assessed. The correlation matrix, constructed across all alert sounds, indicated the presence of multiple coeffi-

TABLE 2
Measures of Central Tendency for the 10 Alert Sounds

	M	*Mode*	*Mdn*	*SD*
Effectiveness	5.17	7.00	6.00	1.87
Familiarity	4.94	7.00	6.00	2.48
Association	5.30	7.00	7.00	2.15
Distinctiveness	5.83	7.00	7.00	1.69
Detection	5.53	7.00	6.00	1.73
Urgency	5.08	7.00	6.00	1.91
Startle	3.92	2.00	4.00	1.92
Irritation	2.75	2.00	2.00	1.52
Distraction	2.91	2.00	3.00	1.55

Note. $N = 486$.

cients of .3 and above. The Kaiser–Meyer–Oklin measures of sampling adequacy were .71 and the Bartlett's Test of Sphericity reached significance ($p < .001$), supporting the factorability of the correlation matrix.

The scree plot revealed the presence of two components. Likewise, the PCA indicated the presence of two components with eigenvalues exceeding 1.0, explaining 56.7% of the total variance (Component 1 = 34.1%, Component 2 = 22.6%). To aid interpretation, a varimax rotation was performed on the component matrix (Table 3). Loadings below 0.4 were suppressed in the analysis.

The first component was assigned the title "alert information" because distinctiveness, familiarity, association, and ease of detection variables all related to the content of the alert. Variables in Component 1 significantly correlated with alert effectiveness ratings and each other. In other words, alerts that received high ratings on distinctiveness, familiarity, association, and ease of detection also received high effectiveness ratings. Component 2 was assigned the title "alert response," as likelihood of startle, irritation, and distraction are all variables that have been documented to influence pilot performance following alert detection. The variables in Component 2 correlated significantly with each other but were not negatively correlated with effectiveness scores. This suggests that alert content and style of presentation were viewed as separate issues. The pilots appeared to prioritize the alert content over the style of presentation when rating overall effectiveness.

Urgency Mapping

Within the questionnaire all the participants were asked two urgency-related questions. The first asked the pilots to rate the situational urgency if each of the following alerts were triggered: fire, traffic, windshear, system caution, configuration, ground proximity, autopilot disconnect, overspeed, decision height, and 500 ft. The second question asked the pilots to rate the perceived urgency of each of the 10 alert sounds. To assess the level of urgency mapping for the set of alerts, a correlation was performed on the mean event and mean alert scores. The correlation coefficient was .71 (N = 10), which was significant at the .05 level, indicating that, on the whole, according to pilot opinion, the urgency mapping for the alerts was accurate. The scatter plot for this correlation is given in Figure 2.

Accurate urgency mapping can be observed for the majority of the alerts, especially the ground proximity, windshear, and autopilot disconnect alerts. However, urgency mismatching can be observed in the overspeed, decision height, 500 ft, and system caution alerts.

TABLE 3
Rotated Component Matrix

	Component 1: Alert Information	*Component 2: Alert Response*
Effectiveness	.735	—
Association	.813	—
Familiarity	.744	—
Distinctiveness	.675	—
Detection	.623	—
Startle	—	.642
Irritation	—	.777
Distraction	—	.806

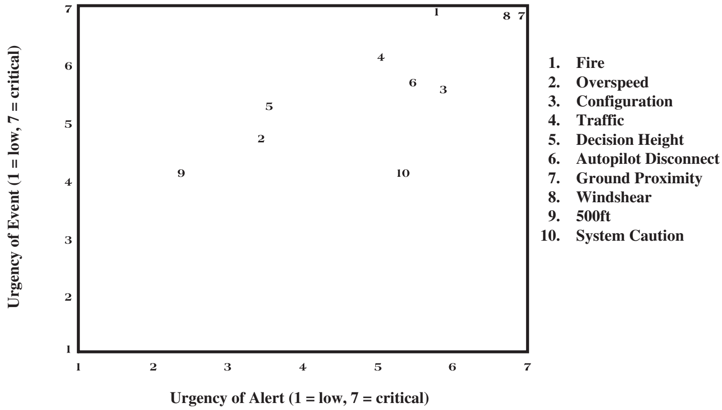


FIGURE 2 Urgency mapping scatter plot.

DISCUSSION

The purpose of this research was to obtain feedback from pilots on current design principles. Section A of the survey assessed the subjective opinions of pilots on current alert intensity levels, and Section B assessed overall effectiveness ratings of a set of auditory alerts currently in operation.

Alert Intensity

The majority of participants tested in the survey currently fly the most modern generation of aircraft and it is interesting to note that despite the vast changes to visual display design on the flight deck, the way auditory alerts are presented to pilots has remained much the same. The criticisms referred to in Patterson's (1982) report more than 20 years ago are remarkably similar to the criticisms identified in this study. Current auditory alerts perform their function well but:

> [A]chieve their success at considerable cost, in that they flood the flight deck with very loud, strident sounds. This has two unfortunate side effects: first, it makes the auditory warning system unpopular with the flight crew. Secondly, and perhaps more important, many of the existing warnings disrupt thought and prevent crew communication, which at a critical moment makes an already difficult situation worse. (Patterson, 1982, p. 1)

More than half of the pilots rated the current intensity level of alerts as too high. Nearly three quarters of the pilots tested had experienced startle from auditory alerts, and approximately half of the participants had also experienced impaired performance due to auditory alert presentation. New technological advances in adding active noise-reduction capabilities to pilot headsets have further exacerbated the problem of loud alerts. Active noise-reduction headsets typically reduce flight deck noise levels by approximately 10 to 15dB, which further increases the perception of alert loudness and likelihood of startle.

Alert Preferences

The majority of pilots tested responded in favor of combination alerts. Combination alerts contain a nonspeech prealert that is designed to capture attention and a speech component that verifies alert meaning. For example, "I feel that an alarm sound is good at getting attention and the voice that follows is an unambiguous description of the fault or the action to take. Voice on its own is too easy to disregard in high-workload situations and bells and sirens are too confusing on their own." An example of a combination alert evaluated in this survey is the ground proximity alert commonly referred to as "Whoop, Whoop—Pull Up" (see Appendix).

Alert Ratings

Taken together, the results suggest that the participants rated the design and informational content of the alerts favorably. This is indicated by the high scores on the distinctiveness, association, and ease of detection variables; however, for many of the alerts there is a likelihood of startle, and if presented repetitively or continuously, a likelihood of irritation and distraction (see Appendix). A startle response is mainly triggered due to a large change in sound pressure level. If an alert reaches its peak intensity instantaneously, there is a greater potential for startle. A method of reducing the startling quality of the alerts would be to reassess the onset intensity. Assigning a gradient of intensity increase is argued to reduce the chance of startle (Patterson, 1982). Distraction and irritation, on the other hand, are more associated with the duration of an alert. To guarantee detection, alerts need to have distracting qualities; however, if they continue after the point of detection, these distracting qualities can interfere with resolution of the situation (Banbury et al., 2001; Wickens & Hollands, 2000).

In the PCA, variables relating to the design and informational content of the alert (Component 1) all had strong positive correlations with effectiveness, whereas the variables that related to the style of alert presentation did not negatively correlate with effectiveness and loaded onto a separate component. Despite unfavorable scores on startle, distraction, and irritation variables, many of the alerts received high effectiveness ratings nonetheless (see Appendix). The dichotomy between components suggests that pilots regard style of presentation and sound content as distinct issues when assessing alert efficacy.

Urgency Mapping

The majority of the alerts received accurate urgency mapping, although there are two alerts that are of concern. First, the acoustic characteristics of the system caution alert, commonly referred to as the "owl," were rated as more urgent than the event dictates. Particular attention needs to be applied to caution alerts so that tasks of higher priority are not interrupted by less urgent events. A way of decreasing the perceived urgency of this alert would be to either lower the pitch or increase the interpulse interval. Second, the overspeed "clacker" was perceived to be far less urgent than the event dictates. The perceived urgency of this alert could be increased by either raising the pitch or decreasing the interpulse interval. One caveat associated with this analysis is that it is unclear if pilots used previous experience when rating the perceived urgency of the alert sounds.

To ensure that urgency is assessed on acoustic characteristics alone, future urgency mapping analyses will include a group of novice listeners.

Future Directions in Auditory Alert Presentation

To improve auditory alert assistance, designers should attempt to find methods of lowering alert intensity without compromising detection. Patterson (1982) suggested presenting alerts at a fixed ratio above flight deck noise. However, one limitation of Patterson's procedure is that presentation levels are set using the highest levels of routine flight noise. As Edworthy and Hellier (2000) pointed out, "both the level and spectrum of aircraft noise will vary as a function of speed, height, and current activity, [and] these variables need to be taken into account in some way" (p. 29).

In an attempt to build on Patterson's (1982) ideas, research at the University of Bristol has applied new technological advances to controlling the intensity of auditory alerts. A possible solution to excessively loud alerts is to conduct contemporaneous analysis of flight deck noise. A system that is currently being evaluated samples flight deck noise in real time via a microphone in the pilot's headset (for headset presentation) or headrest (for speaker presentation). Amplitude and spectral characteristics of the noise are analyzed in software. The output of this procedure is fed into additional software that is responsible for attenuating or augmenting the alerts to the desired level. After the alert intensity has been set, the alert is then delivered to the pilot via the headset or flight deck speakers. There are two advantages in employing this system. First, alert intensity levels model background noise levels at the time of presentation. As a result, alerts are presented at a lower level during quieter phases of flight and the potential for startle and communications masking is reduced. Future designs of this system aim to incorporate active noise reduction, allowing additional control over alert intensity. Second, the alert presentation software can be automated to perform more complex presentation patterns. For example, the sound can be lowered to 5dB above the detection threshold once the alert has been acknowledged. At this level the signal will still be audible but should not dominate pilot attention. This is hypothesized to offer a more suitable environment for decision making, problem solving, and communication during warning events. Clearly, examination of such modifications requires scientific research within the flight deck environment to ensure efficacy and suitability.

APPLICATIONS FOR THIS RESEARCH

The research findings reported here show that current auditory alert designs are prone to the same criticisms that were identified more than 20 years ago. This suggests that there has been no resolution to the problems of excessive loudness and duration of alert messages. Startle, communications masking, and task distraction increase the potential for pilot error during warning events and make the alert systems unpopular with flight crews. It is therefore important that future research continues to examine the presentation style of alert messages to discover more efficient ways of using auditory signals in warning events.

ACKNOWLEDGMENTS

This research was funded by the Civil Aviation Authority (U.K.). We would also like to acknowledge British Airways and the Bluecoat Organization for supplying participants, and Charles Jenkins for technical help with the survey.

REFERENCES

Arrabito, G. A., Mondor, T. A., & Kent, K. J. (2004). Judging the urgency of non-verbal auditory alarms: A case study. *Ergonomics, 47,* 821–840.

Banbury, S. P., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors, 43,* 12–29.

Björk, E. A. (1999). Startle, annoyance and psychophysical responses to repeated sound bursts. *Acustica (Acta Acustica), 85,* 575–578.

Bliss, J. P. (2003). Investigation of alarm-related accidents and incidents in aviation. *International Journal of Aviation Psychology, 13,* 249–268.

Bradley, M. M., & Lang, P. L. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37,* 204–215.

Burt, J. L., Bartolome, D. S., Burdette, D. W., & Comstock, J. R. (1995). A psychophysical evaluation of the perceived urgency of auditory warning signals. *Ergonomics, 38,* 2327–2340.

Doll, T. J., & Folds, D. J. (1986). Auditory signals in military aircraft: Ergonomic principles versus practice. *Applied Ergonomics, 17,* 257–264.

Edworthy, J. (1994). Urgency mapping in auditory warning signals. In N. Stanton (Ed.), *Human factors in alarm design* (pp. 15–30). London: Taylor & Francis.

Edworthy, J., & Hellier, E. (2000). Auditory warnings in noisy environments. *Noise and Health, 6,* 27–39.

Edworthy, J., Hellier, E., Walters, K., Weedon, B., & Adams, A. (2000). The relationship between task performance, reaction time, and perceived urgency in nonverbal auditory warnings. In *Proceedings of IEA 2000, HFES 2000 Congress 3* (pp. 674–677). San Diego, CA: Human Factors and Ergonomic Society.

FAA Advisory Circular. (2000). *Certification of transport airplane mechanical systems* (AC 25–22). Washington, DC: Federal Aviation Administration.

Grillon, C., Pellowski, M., Merikangas, K. R., & Davis, M. (1997). Darkness facilitates the acoustic startle reflex in humans. *Biological Psychiatry, 42,* 453–460.

Haas, E. C., & Casali, J. G. (1995). Perceived urgency of and response-time to multitone and frequency modulated warning signals in broadband noise. *Ergonomics, 38,* 2313–2326.

Haas, E. C., & Edworthy, J. (1996). Designing urgency into auditory warnings using pitch, speed and loudness. *Computing and Control Engineering Journal, 7,* 193–198.

Hawkins, F. H. (1993). *Human factors in flight* (2nd ed.). Aldershot, England: Ashgate.

Hellier, E., & Edworthy, J. (1999). On using psychophysical techniques to achieve urgency mapping in auditory warnings. *Applied Ergonomics, 30,* 167–171.

INT DEF STAN 00–25, Part 19/1. (2004). *Human engineering domain.* Glasgow, Scotland: Defence Procurement Agency.

ISO 7731. (1986). *Danger signals for work places—Auditory danger signals.* Geneva, Switzerland: Organization for International Standardization.

James, S. H. (1996, October). *Audio warnings for military aircraft.* Paper presented at the AGARD Aerospace Medical Panel Symposium on Audio Effectiveness in Aviation, Copenhagen, Denmark.

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2001, March). *Cockpit interruptions and distractions: A line observations study.* Paper presented at the 11th International Symposium on Avaition Psychology, Columbus, OH.

Mårtensson, L. (1995). The aircraft crash at Gottröra: Experiences of the cockpit crew. *International Journal of Aviation Psychology, 5,* 305–326.

MIL–STD–1427F. (1999). *Department of Defense design criteria standard: Human engineering.* Washington, DC: U.S. Department of Defense.

Momtahan, K., Hétu, R., & Tansley, B. (1993). Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics, 36,* 1159–1176.

National Safety Data Analysis Center. (1993). Brief Report No. 260299 [Online]. Retrieved May 2005, from https://extranet.nasdac.faa.gov/pls/portal/STAGE.ASRS_BRIEF_REPORT?RPT_NBR=260299&AC_VAR=TRUE&RPRT_VAR=TRUE&ANMLY_VAR=TRUE&SYN_VAR=TRUE&NARR_VAR=TRUE&NARR_SRCH=

National Safety Data Analysis Center. (1998). Brief Report No. 412066 [Online]. Retrieved May 2005, from https://extranet.nasdac.faa.gov/pls/portal/STAGE.ASRS_BRIEF_REPORT?RPT_NBR=412066&AC_VAR=TRUE&RPRT_VAR=TRUE&ANMLY_VAR=TRUE&SYN_VAR=TRUE&NARR_VAR=TRUE&NARR_SRCH=

National Safety Data Analysis Center. (2002). Brief Report No. 543414 [Online]. Retrieved May 2005, from https://extranet.nasdac.faa.gov/pls/portal/STAGE.ASRS_BRIEF_REPORT?RPT_NBR=543414&AC_VAR=TRUE&RPRT_VAR=TRUE&ANMLY_VAR=TRUE&SYN_VAR=TRUE&NARR_VAR=TRUE&NARR_SRCH=

Noyes, J. M., & Starr, A. F. (2000). Civil aircraft warning systems: Future directions in information management and presentation. *International Journal of Aviation Psychology, 10,* 169–188.

Patterson, R. D. (1982). *Guidelines for auditory warnings on civil aircraft* (CAA Paper No. 82017). London: Civil Aviation Authority.

Pritchett, A. R. (2001). Reviewing the role of cockpit alerting systems. *Human Factors and Aerospace Safety, 1,* 5–39.

Rigby, L. V., & Edelman, D. A. (1968). A predictive scale of aircraft emergencies. *Human Factors, 10,* 475–482.

Sorkin, R. D. (1988). Why are people turning off our alarms? *Journal of the Acoustical Society of America, 84,* 1107–1108.

Thackray, R. I., & Touchstone, R. M. (1970). Recovery of motor performance following startle. *Perception and Motor Skills, 30,* 279–292.

Turpin, G., Schaefer, F., & Boucsein, W. (1999). Effects of stimulus intensity, risetime, and duration on autonomic and behavioral responding: Implications for the differentiation of orienting, startle, and defense responses. *Psychophysiology, 36,* 453–463.

Uchtdorf, D., & Heldt, P. (1989, April). Survey on cockpit systems B737–300/ A310–200, 1986 [Special issue]. *Flightcrew Info.*

Ulfvengren, P. (2003). Associability: A comparison of sounds in a cognitive approach to auditory alert design. *Human Factors and Aerospace Safety, 3,* 313–331.

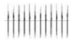
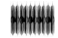
Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.) Upper Saddle River, NJ: Prentice Hall.

Wilkins, P. A. (1982). Effects of noise on people. In R. G. White & J. G. Walker (Eds.), *Noise and vibration* (pp. 805–825). New York: Wiley.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics, 38,* 2371–2393.

Manuscript First Received: December 2003

APPENDIX
Mean Ratings for Each of the 10 Alerts

	Fire	*Overspeed*	*Configuration*	*Traffic*	*Decision Height*	*Autopilot Disconnect*	*Ground Proximity*	*Windshear*	*500 ft*	*System Caution*
										
Effectiveness										
M	5.5	4.1	5.3	4.9	3.2	5.3	6.2	6.5	5.6	4.8
SD	1.8	1.6	1.7	1.9	1.5	1.7	1.7	1.3	1.7	1.7
Familiarity										
M	5.4	3.0	5.2	6.3	1.7	4.3	6.7	6.5	6.3	3.8
SD	2.3	2.5	2.2	1.7	1.6	2.2	0.9	1.2	1.6	1.6
Distinctiveness										
M	5.8	3.9	4.7	6.2	2.8	4.9	6.7	6.5	6.4	3.5
SD	1.7	1.7	2.1	1.8	1.8	2.1	1.2	1.5	1.4	1.4
Detection										
M	6.0	5.5	5.7	6.6	4.7	5.6	6.6	6.6	5.9	5.0
SD	1.6	1.7	1.7	1.2	1.9	1.5	1.2	1.3	1.8	1.8
Association										
M	5.6	4.3	6.2	5.8	3.9	5.9	6.6	6.7	4.5	5.3
SD	1.6	1.7	0.8	1.6	1.7	1.2	1.3	0.8	1.9	1.9
Urgency										
M	5.8	3.4	5.9	5.0	3.5	5.5	6.9	6.7	2.5	5.4
SD	1.8	1.5	1.1	1.6	1.5	1.5	0.3	0.5	1.5	1.5
Startle										
M	3.6	4.7	3.5	4.4	4.7	3.4	2.3	2.8	6.1	3.8
SD	1.9	1.6	1.8	2.0	1.2	1.7	1.5	1.8	1.0	1.0
Irritation										
M	2.2	2.9	2.1	3.8	3.0	2.3	2.8	2.8	3.3	2.4
SD	1.5	1.5	1.4	1.5	1.2	1.1	1.5	1.5	2.0	2.0
Distraction										
M	2.7	3.2	2.4	3.7	3.0	2.4	3.1	3.0	3.4	2.5
SD	1.8	1.5	1.6	1.4	1.4	1.3	1.6	1.4	1.7	1.7